

Rec'd PCT/PTO 22 FEB 2003

10/501857  
PCT/EP US / 00005



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

REC'D 27 FEB 2003  
WIPO PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-  
gen stimmen mit der  
ursprünglich eingereichten  
Fassung der auf dem näch-  
sten Blatt bezeichneten  
europäischen Patentanmel-  
dung überein.

The attached documents  
are exact copies of the  
European patent application  
described on the following  
page, as originally filed.

Les documents fixés à  
cette attestation sont  
conformes à la version  
initialement déposée de  
la demande de brevet  
européen spécifiée à la  
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02001255.5

**PRIORITY  
DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1 (a) OR (b)

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

R C van Dijk



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

**Blatt 2 der Bescheinigung**  
**Sheet 2 of the certificate**  
**Page 2 de l'attestation**

Anmeldung Nr.:  
Application no.: 02001255.5  
Demande n°:

Anmeldetag:  
Date of filing: 17/01/02  
Date de dépôt:

Anmelder:  
Applicant(s):  
Demandeur(s):  
SIEMENS AKTIENGESELLSCHAFT  
80333 München  
GERMANY

Bezeichnung der Erfindung:  
Title of the invention:  
Titre de l'invention:

Verfahren zur einfachen Erweiterung eines Spracherkennungsvokabulars

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:  
State:  
Pays:

Tag:  
Date:  
Date:

Aktenzeichen:  
File no.  
Numéro de dépôt:

Internationale Patentklassifikation:  
International Patent classification:  
Classification internationale des brevets:  
G10L15/26

Am Anmeldetag benannte Vertragsstaaten:  
Contracting states designated at date of filing:  
Etats contractants désignés lors du dépôt:

AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR

Bemerkungen:  
Remarks:  
Remarques:

Ursprüngliche Bezeichnung der Anmeldung: Siehe bitte Seite 1 der Beschreibung .

17. Jan. 2002

## Beschreibung

## Spracherkenner und Betriebsverfahren für einen solchen

- 5 Die Erfindung betrifft einen Spracherkenner nach dem Oberbegriff des Anspruchs 1 sowie ein Betriebsverfahren für einen solchen.

10 Nachdem sich die Spracherkennung bei der Texteingabe in auf PCs laufenden Büroapplikationen seit Jahren ein festes und stetig wachsendes Anwendungsgebiet gesichert hat, dringt sie zunehmend auch in die Steuerung technischer Geräte ein. So-

15 wohl bei hochgradig miniaturisierten und zugleich computerisierten handgehaltenen elektronischen Geräten - insbesondere Mobiltelefonen und PDAs - als auch bei technischen Geräten, deren Bedienung möglichst wenig Aufmerksamkeit und Konzentration des Bedieners binden sollen - wie der verschiedenen technischen Geräte in einem fahrenden Auto - kann diese Art der Spracherkennung, mit darauf aufbauender Sprachsteuerung,

20 sinnvolle Einsatzmöglichkeiten finden. Bei der erstgenannten Art von Geräten ist nämlich die für Bedienvorgänge verfügbare Fläche so klein geworden, daß sich die Vielzahl möglicher Funktionen durch herkömmliche Tastatur- oder TouchScreen-Eingaben nur noch sehr unbequem (und für Personen mit

25 schlechtem Sehvermögen nahezu überhaupt nicht mehr) realisieren läßt. Bei Einsatzfeldern, in denen die Aufmerksamkeit des Benutzers auf andere Dinge (z. B. den Straßenverkehr) gerichtet bleiben muß, erbringt die Einführung der Sprachsteuerung neben der Komfortsteigerung auch einen wesentlichen Sicherheitszuwachs.

30

In der Spracherkennung wird ein Lexikon mit den zu erkennen-

35 den Wörtern benötigt. Diese werden im Falle der phonembasier-ten Spracherkennung mittels einem Text-to-Phonem-Verfahren in eine phonetische Umschrift überführt und im Vokabular abgelegt. Während des Erkennungsvorgangs erfolgt im sogenannten Viterbi-Algorithmus eine Suche nach dem besten Pfad durch die

im Vokabular enthaltenen Phonemfolgen. Einzelheiten zu den etablierten Spracherkennungsalgorithmen finden sich in der einschlägigen Fachliteratur.

5 Weitgehend computerisierte technische Geräte der oben angesprochenen Art (PDAs, Handheld-PCs, Mobiltelefone, Kfz-Audioanlagen, Kfz-Borcomputer etc.) haben vielfach an PC-Bedienoberflächen angelehnte Strukturen der Benutzerschnittstellen bzw. MMIs. Es ist eine Mehrzahl von Applikationen instal-

10 liert, die auf geeignete Weise - und bei komplexeren Geräten auch in einer bestimmten, logischen hierarchischen Unterordnung - angesteuert werden müssen. Bei herkömmlichen Geräten dieser Art wird hierzu eine Menüsteuerung angeboten, die vom Benutzer durch Softkey-Eingaben ausgeführt wird.

15 Bei der Applikationswahl per Spracheingabe befinden sich im Lexikon die Programmnamen der verfügbaren Applikationen. Nach der Erkennung eines Namens wird das jeweilige Programm ausgeführt bzw. die Applikation gestartet. Dazu müssen die Programmnamen und die Programmpfade in einem geeigneten Format

20 abgespeichert sein.

Gemäß dem Stand der Technik werden die einzelnen Programmnamen mit den entsprechenden Erkennungsergebnissen (den Worten

25 des Lexikons) fest verdrahtet. Dies kann einerseits in einer zusätzlichen Datei festgelegt werden oder andererseits fest im Quellcode des Programms definiert sein. Beide Methoden haben entscheidende Nachteile, welche im folgenden beschrieben werden:

30 - Beim Arbeiten mit einer zusätzlichen Datei besteht das Problem, daß diese von einem Benutzer einsehbar ist und demzufolge auch verändert werden kann. Auch binäre Formate oder schreibgeschützte Dateien bieten keinen wirksamen Schutz vor

35 Veränderungen. Dadurch können Diskrepanzen zwischen dem verwendeten Vokabular und der Wortliste bzw. Programmliste ent-

stehen. Dies hat auch zur Folge, daß die Applikation möglicherweise falsch reagiert.

- 5 - Bei der Festlegung der als Steuerbefehle wirkenden sprachlichen Äußerungen im Quellcode kann das Vokabular auf einfachem Wege nicht weiter geändert werden. Es müßte der Quellcode jedes Mal neu übersetzt und ausgeliefert werden, sobald Änderungen der Programmnamen auftreten.
- 10 - Der entscheidende Nachteil des bisherigen Vorgehens liegt in der fehlenden bzw. unzureichenden Erweiterbarkeit des Systems. Bisher ist es dem Benutzer nicht möglich, seine eigenen Kommandos oder Anwendungen in die automatische Spracherkennung mit aufzunehmen - jedenfalls nicht ohne die Gefahr einer
- 15 Störung der ursprünglich programmierten Konfiguration des Spracherkenners.

Der Erfindung liegt daher die Aufgabe der Bereitstellung eines verbesserten Spracherkenners und Verfahrens zum Betrieb

20 eines solchen zugrunde, mit denen dieser zur Berücksichtigung von eigenen Steuerbefehlen oder Anwendungen des Nutzers flexibler konfiguriert werden kann.

Diese Aufgabe wird in ihrem Vorrichtungsaspekt durch einen

25 Spracherkenner mit den Merkmalen des Anspruchs 1 und in ihrem Verfahrensaspekt durch ein Betriebsverfahren mit den Merkmalen des Anspruchs 6 gelöst.

Die Erfindung schließt den grundlegenden Gedanken ein, zur

30 Sprachsteuerung von Applikationen - bzw. zu einer entsprechenden Handhabung von Dateien - eine mit Links aufgebaute Benutzeroberfläche bereitzustellen. Das Organisationsprinzip der Links ermöglicht ohne weiteres einen strukturierten Aufruf von Programmen bzw. Dateien in unterschiedlichen Hierarchieebenen, ohne daß a priori eine starre Zuordnung definiert

35 und programmiert werden muß.

- Die Liste der zu erkennenden Wörter (das Lexikon) wird durch den Inhalt eines bestimmten Dateiverzeichnisses bestimmt, in welchem Links (Verknüpfungen) zu den vorhandenen Programmen bzw. Dateien enthalten sind. Der Name des Links gibt das zu
- 5 erkennende Wort an, und das Programm bzw. die Datei, auf die dieser Link zeigt, gibt die auszuführende Aktion an. Bei der Umsetzung des Namens sollte darauf geachtet werden, daß nur der Teilstring vor dem ersten Punkt als Kommando verwendet
- 10 wird. Beim Start des Erkennerprogramms wird das Vokabular erzeugt. So kann flexibel auf Änderungen der Applikations- bzw. Dateistruktur reagiert werden. Sobald ein Wort erkannt wird, wird der entsprechende Link aktiviert und die gewünschte Aktion ausgeführt.
- 15 Vorteile gegenüber dem bisherigen Verfahren sind in der Flexibilität, was Worte und Aktionen betrifft, und der einfachen Erstellung und Veränderung eines komplexen Erkennerwortschatzes zu sehen. Das Hinzufügen von neuen Kommandos zum bisherigen Wortschatz kann einfach und in gewohnter Weise durchge-
- 20 führt werden. Im Dateiverzeichnis muß lediglich eine Verknüpfung zum gewünschten Programm oder der Datei erstellt werden. Unter Windows z. B. ist das Erstellen eines Links einfach über das Kontextmenü möglich.
- 25 Damit wird ein weiterer Vorteil klar: Da das Filesystem die Verwaltung von Kommandos und Aktionen (Name und Ziel der Verknüpfung) übernimmt, ist kein zusätzliches Programm zur Verwaltung des Wortschatzes nötig. Soll ein Kommando gelöscht werden, wird einfach der Link gelöscht.
- 30 Da moderne Betriebssysteme auch Links auf Dateien erlauben, können Dokumente ebenfalls per Sprachkommando geöffnet werden.
- 35 In einer bevorzugten Ausführung umfaßt das Dateiverzeichnis eine Mehrzahl von Unterverzeichnissen in mindestens einer untergeordneten Hierarchieebene, wobei die Verzeichnisnamen ein

erstes und gegebenenfalls weitere, hierarchisch untergeordnete aktive Teilvokabulare des Spracherkenners bilden.

Durch die Verwendung von Unterverzeichnissen im Dateiverzeichnis lassen sich auf einfachste Weise strukturierte Sprachaufrufe von Programmen und Dateien erzeugen. So können z. B. alle Links zu Musikstücken in ein Unterverzeichnis "Musik" abgelegt werden. In der ersten Stufe der Erkennung befindet sich das Wort "Musik" im aktiven Vokabular. Wird dieses erkannt, wird das Vokabular gewechselt (z. B. per Language Model), die im Unterverzeichnis "Musik" enthaltenen Links stehen nun im aktiven Vokabular.

Insbesondere ist jedem Programm bzw. jeder Datei aus einem Unterverzeichnis ein zusammenhängender, mehrgliedriger Sprachbefehl zugeordnet, der die Namen der zu dem Programm bzw. der Datei führenden Links des Dateiverzeichnisses und jedes nachgeordneten Unterverzeichnisses umfaßt.

Mit dieser Methode lassen sich in einfachster Weise komplexe Sprachkommandos erstellen und editieren. Vorhandene Verzeichnisse mit Verknüpfungen, wie z. B. das Windows Startmenü, lassen sich nun einfach per Sprachsteuerung bedienen, da schon alle nötigen Informationen vorhanden sind.

Dieses Verfahren ist eine Weiterentwicklung von Verknüpfungen mit Programmen (Beispiel Windows PC) und den festverdrahteten Spracherkennerressourcen. Hierbei wird durch Anlegen eines Links automatisch die Erkennerressource erstellt. D. h. danach kann sofort der Name des Links durch den Erkennen verarbeitet werden.

Generell können beliebige Dateien und Programme, nachdem sie in das spezielle Verzeichnis kopiert wurden, per Sprachbefehl gestartet werden. Es ist also egal, ob es sich um einen Musiktitel, eine c++Datei, ein Textdokument oder um ein Programm handelt. Durch das Abspeichern eines Links in dem spe-

ziellen Verzeichnis wird die Datei durch das eingestellte Defaultprogramm geöffnet. Beispiel: Ein Dokument mit der Endung .doc wird automatisch mit dem Programm Word geöffnet (wie beim Doppelklick mit der Maus auf die Datei bei einer herkömmlichen PC-Eingabe).

Die oben erläuterten Aspekte der Erfindung stellen sich sowohl als Vorrichtungsaspekt eines Spracherkenners wie auch als Aspekte des Betriebsverfahrens eines solchen dar - zumal die Realisierung typischerweise in einer geeigneten Mischung aus Hard- und Softwarekomponenten erfolgen wird.

Nachfolgend sind zwei Arten, ein Wort in das Erkennerlexikon aufzunehmen, genannt:

(1) Aufnahme durch einen Programmaufruf über das Kontextmenü für die gewünschte Anwendung. Hierbei enthält das Kontextmenü zwei Programmaufrufe (Add und Remove). Add fügt das entsprechende Programm/die Datei hinzu und Remove zeigt die Liste der Programme/Dateien an, welche aktuell durch Sprachwahl ausgewählt werden können.

(2) Kopieren des Link der gewünschten Anwendung in den speziellen Ordner über "drag'n'drop". Hierbei muß, um ein Programm zu entfernen, in das entsprechende Verzeichnis gewechselt und der gewünschte Link durch "Entfernen" aus dem Verzeichnis gelöscht werden.

Die Ausführung der Erfindung ist nicht auf die oben beschriebenen Beispiele und Aspekte beschränkt, sondern ebenso in einer Vielzahl von Abwandlungen möglich, die im Rahmen fachgemäßen Handelns liegen.



17. Jan. 2002

## Patentansprüche

1. Spracherkenner mit einem gespeicherten Vokabular von zu  
erkennenden Wörtern zur Sprachsteuerung einer Mehrzahl von  
5 Programmen und/oder sonstigen Dateien, denen jeweils ein Wort  
des Vokabulars als Name zugeordnet ist,  
d a d u r c h g e k e n n z e i c h n e t, daß  
in einem Dateiverzeichnis zu jedem Programm bzw. jeder Datei  
ein Link gespeichert ist, wobei die Namen der Links ein ers-  
10 tes aktives Teilvokabular des Spracherkenners bilden.
2. Spracherkenner nach Anspruch 1,  
d a d u r c h g e k e n n z e i c h n e t, daß  
die Namen der Links durch Sprachbefehle gebildet sind und die  
15 Links Verknüpfungen zu Applikationsprogrammen definieren.
3. Spracherkenner nach Anspruch 1,  
d a d u r c h g e k e n n z e i c h n e t, daß  
die Namen der Links durch Sprachbefehle gebildet sind und die  
20 Links Verknüpfungen zu Dokumenten, insbesondere Textdokumen-  
ten oder Sprach-, Musik- oder Videodateien, definieren.
4. Spracherkenner nach einem der vorangehenden Ansprüche,  
d a d u r c h g e k e n n z e i c h n e t, daß  
25 das Dateiverzeichnis eine Mehrzahl von Unterverzeichnissen in  
mindestens einer untergeordneten Hierarchieebene umfaßt, wo-  
bei die Namen der Unterverzeichnisse zusammen mit denen der  
Links ein erstes und gegebenenfalls weitere, hierarchisch un-  
tergeordnete aktive Teilvokabulare des Spracherkenners bil-  
30 den.
5. Spracherkenner nach Anspruch 4,  
d a d u r c h g e k e n n z e i c h n e t, daß  
jedem Programm bzw. jeder Datei aus einem Unterverzeichnis  
35 ein zusammenhängender, mehrgliedriger Sprachbefehl zugeordnet  
ist, der die Namen der zu dem Programm bzw. der Datei führen-

den Links des Dateiverzeichnisses und jedes nachgeordneten Unterverzeichnisses umfaßt.

5 6. Betriebsverfahren eines Spracherkenners nach einem der vorangehenden Ansprüche,  
d a d u r c h g e k e n n z e i c h n e t, daß  
beim Start des Spracherkennerprogramms das aktuelle Vokabular erzeugt wird, welches wenigstens die Namen der Links des Dateiverzeichnisses umfaßt.

10 7. Betriebsverfahren nach Anspruch 6,  
d a d u r c h g e k e n n z e i c h n e t, daß  
die Administrierung des Vokabulars als Verwaltung des Dateiverzeichnisses und optional vorhandener Unterverzeichnisse,  
15 ohne zusätzliches Vokabular-Verwaltungsprogramm, erfolgt.

8. Betriebsverfahren nach Anspruch 6 oder 7,  
d a d u r c h g e k e n n z e i c h n e t, daß  
zur Editierung zusammenhängender, mehrgliedriger Sprachbefeh-  
20 le unter dem Dateiverzeichnis Unterverzeichnisse in mindestens einer untergeordneten Hierarchieebene angelegt werden  
und die Erkennung zusammenhängender, mehrgliedriger Sprachbefehle in einem mehrstufigen Erkennungsvorgang erfolgt, in  
dessen Ablauf von einem ersten in ein zweites und gegebenen-  
25 falls weitere aktive Teilvokabulare gewechselt wird.

9. Betriebsverfahren nach einem der Ansprüche 6 bis 8,  
d a d u r c h g e k e n n z e i c h n e t, daß  
die Aufnahme neuer Worte in das Vokabular oder die Entfernung  
30 von Worten aus dem Vokabular durch einen Programmaufruf über ein an sich bekanntes Kontextmenü für das betreffende Programm bzw. die betreffende Datei oder durch einen  
"drag'n'drop"-Ablauf erfolgt.

EPO - Munich  
61  
17. Jan. 2002

## Zusammenfassung

### Spracherkenner und Betriebsverfahren für einen solchen

- 5    Spracherkenner mit einem gespeicherten Vokabular von zu er-  
kennenden Wörtern zur Sprachsteuerung einer Mehrzahl von Pro-  
grammen und/oder sonstigen Dateien, denen jeweils ein Wort  
des Vokabulars als Name zugeordnet ist, wobei in einem Datei-  
verzeichnis zu jedem Programm bzw. jeder Datei ein Link ge-  
speichert ist, wobei die Namen der Links ein erstes aktives  
10    Teilvokabular des Spracherkenners bilden.